



# Object recognition in hyperspectral images using Binary Partition Tree representation

Silvia Valero<sup>a</sup>, Philippe Salembier<sup>b</sup>, Jocelyn Chanussot<sup>c,d</sup>

<sup>a</sup>CESBIO - CNES, CNRS (UMR 5126), IRD, Université de Toulouse, France

<sup>b</sup>Technical University of Catalonia (UPC), Barcelona, Catalonia, Spain

<sup>c</sup>Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland

<sup>d</sup>GIPSA-lab, Signal & Image Dept., Grenoble Institute of Technology, Grenoble, France

## ARTICLE INFO

### Article history:

Object based image analysis  
Binary Partition Tree  
Hyperspectral Images  
Object Recognition

## ABSTRACT

In this work, an image representation based on Binary Partition Tree is proposed for object detection in hyperspectral images. This hierarchical region-based representation can be interpreted as a set of hierarchical regions stored in a tree structure, which succeeds in presenting: (i) the decomposition of the image in terms of coherent regions and (ii) the inclusion relations of the regions in the scene. Hence, the BPT representation defines a search space for constructing a robust object identification scheme. Spatial and spectral information are integrated in order to analyze hyperspectral images with a region-based perspective. For each region represented in the BPT, spatial and spectral descriptors are computed and the likelihood that they correspond to an instantiation of the object of interest is evaluated. Experimental results demonstrate the good performances of this BPT-based approach.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

Automatic object recognition to map areas has received a lot of attention thanks to the advance of remote sensing technology (Aksoy et al., 2010). In this context, the spatial and the spectral resolutions of the new sensors have played a fundamental role. Specifically, the improvement of the spatial information has been essential for this high-level image understanding task. Accordingly, morphological approaches have received an important interest in gray value or color images (Aytekin and Ulusoy, 2011; Bernstein and Gesu, 1999; Valero et al., 2010).

In the hyperspectral literature, object detection techniques have been mainly developed in the context of pixel-wise spectral classification. In this approach, spectra having a high similarity with the material describing the reference object are individually detected. The drawbacks of pixel-wise analysis is well-known in classical (Cracknell, 1998) and hyperspectral (Fauvel et al., 2012; Tarabalka et al., 2010) remote sensing im-

ages. A major problem is the important semantic gap due to the lack of concordance between the low level and reduced information provided by a single pixel and the human interpretation. Despite this, traditional algorithms characterize objects only by their spectral signatures.

Because of the pixel-based model limitations, research on region-based object detection algorithms has recently received much attention. Region-based representations allow in particular spatial features such as shape, area or orientation to be computed. These features can significantly contribute to the definition of robust object detection algorithms. In this context, the ECognition software (Darwish et al., 2003) was developed. It relies on hierarchical segmentation and produces an image partition on which various region descriptors can be computed. These descriptors are then used as region features for the recognition of objects in the image. One of the main limitations of this strategy is that it assumes that the best partition corresponds to one level of the previously computed hierarchical segmentation. Unfortunately, this assumption is rarely true and, very often, coherent objects can be found at different levels of the hierarchy. Ideally, a robust strategy should study the features in the complete hierarchy to detect the best regions representing the

*e-mail:* [silvia.valero@cesbio.cnes.fr](mailto:silvia.valero@cesbio.cnes.fr) (Silvia Valero),  
[philippe.salembier@upc.edu](mailto:philippe.salembier@upc.edu) (Philippe Salembier),  
[jocelyn.chanussot@gipsa-lab.grenoble-inp.fr](mailto:jocelyn.chanussot@gipsa-lab.grenoble-inp.fr) (Jocelyn Chanussot)

object. As a result, recent works have tried to investigate how different spectral, spatial and joint spectral/spatial features computed on regions evolve from one level to another in a segmentation hierarchy (Plaza and Tilton, 2005). This study proposes to study the regions at different scales, however, no methodology is proposed to automatically select the regions for the objects.

Instead of using a classical hierarchical segmentation approach which produces a single partition, a solution to address the need of multiscale analysis relies on image representations based on regions trees. These representations are useful because beside allowing the study of internal region properties (color, texture, shape, etc.), they also permit the study of external relations such as adjacency, inclusion, similarity of properties, etc. Furthermore, a tree is essentially a hierarchical structure and therefore supports multiscale analysis of regions. The multiscale nature of trees provides flexibility to situations where a given image has to be studied at different scales depending on the processing purpose. In this context, the work presented in (Akay and Aksoy, 2008) proposes the use of component trees resulting from the iterative application of morphological opening and closing on individual PCA spectral bands. The main limitations of this approach are twofold: First, the component tree mainly describes the structure of extrema of the spectral bands and, in hyperspectral images, there is no particular reason why objects of interest should be limited to extrema of spectral bands. Furthermore, in (Akay and Aksoy, 2008), the approach consists in pruning the tree to create a partition before performing the object recognition. The pruning essentially extracts the largest homogenous regions. Once the partition is defined, the search for objects is performed. As in (Horev et al., 2003), one of the drawbacks is that the object detection task is done after a segmentation step producing a partition.

The work presented here proposes to initially generate a hierarchical region-based representation of the image and, then, to use this representation as search space for the object detection (therefore avoiding the creation of a partition for each object) objects are searched as in (Akay and Aksoy, 2008; Horev et al., 2003). A Binary Partition Tree as in (Alonso-Gonzalez et al., 2013) (BPT) is used as hyperspectral image representation. BPTs are less limited than component trees (Akay and Aksoy, 2008) as they do not focus on the description of extrema of spectral bands. They perform a hierarchical grouping based on pixel homogeneity and can directly take into the correlation between spectral bands. For object detection, the use of BPT has been introduced in (Valero et al., 2011) where a simple top-down analysis of the tree branches was done. During this analysis, the objects were detected by selecting the largest nodes having the appropriate features. **Therefore, the detected object was the first node in the branch and the rest of BPT branches were not studied.** However, the best region representing the object is not always the largest one with the appropriate features. Here, we present a more robust strategy that studies all BPT nodes to detect the best ones representing the sought object. The paper organization is as follows: Section 2 introduces the BPT and its construction. The BPT analysis for object detection is discussed in Section 3. Experimental results are reported

BPT is a tree-based structure representing an image by a set of hierarchical regions [15]. In Section 4. Finally, conclusions are drawn in Section 5.

**2. BPT construction**  
The initial partition involves  $n$  regions, a BPT generates a structure containing  $2n-1$  nodes. The BPT is a structured representation of a set of hierarchical partitions which is usually obtained through an iterative bottom-up region merging algorithm [57] [58]. In a bottom-up strategy starting from the leaf nodes, the tree construction is then performed by keeping track of the merging steps. At each step, two pixels or any other initial partition, the tree is constructed by iteratively merging the pair of most similar neighboring regions. Each iteration requires three different tasks: 1) the pair of most similar neighboring regions is identified, 2) a new region corresponding to the union of the region pair is formed, 3) the distance between the new formed region and its neighboring regions is updated. Figure 3.1 illustrates the first iteration  $R_1$  and  $R_4$ . After this iteration, the new formed region  $R_5$  with its neighboring regions is updated. Figure 3.2 shows the pixel example of BPT construction created from an initial partition involving 4 regions.

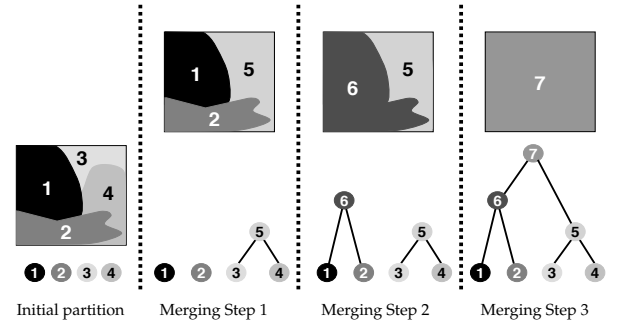


Figure 3.1: Example of BPT construction using a region merging algorithm

In this last figure, the tree leaves corresponds to the regions belonging to the initial partition. However, in order to preserve the resolution in this work, the merging criterion that defines the similarity between a pair of neighboring regions; and 2) a region model that determines how to represent a region and the union of two regions. The BPT construction and, in particular, the region model and the merging criterion have been previously studied for hyperspectral data in Valero et al. (2011, 2013); Alonso-Gonzalez et al. (2013). The region model used here corresponds to the set of normalized histograms of the pixels belonging to each region for each spectral band. Note that this region model based on non parametric probability density functions is used to construct the BPT for texture homogeneity Calderero and Marqués (2010). Using this model with an hyperspectral image containing  $\{\lambda_1, \lambda_2, \dots, \lambda_N\}$  bands, regions are modeled as  $N$  arbitrary discrete distributions, directly estimated from the pixel values.

$$M_R = \{H_R^{\lambda_1}, H_R^{\lambda_2}, \dots, H_R^{\lambda_N}\} \quad (1)$$

$M_R$  is a matrix where each cell represents the probability of the region pixels to have a radiance value in a specific band  $\lambda_k$ . The region model is formed by the rows of the matrix  $H_R^{\lambda_k}$ . It corresponds to the empirical spatial distribution (histogram) of the region  $R$  in the band  $\lambda_k$ .

Note that this model can also be defined when tree leaves are single pixels by using the image self-similarity as in (Dimiccoli and Salembier, 2009; Alonso-Gonzalez et al., 2013). Concerning the merging criterion used to construct the BPT, the criterion proposed in (Cuadras et al., 2012; Valero et al., 2011)

is used here. It relies on distances between observations and canonical correlations and is computed in two steps.

The first step corresponds to a local dimensionality reduction through an analysis of the inter-waveband similarity relationships for each region model. The goal is to remove the redundant hyperspectral information via multidimensional scaling (MDS) (Cox and Cox, 1994), which represents a set of objects as a set of points in a map of chosen dimensionality, based on their interpoint distances. Thus, MDS attempts to locate  $n$  objects as points in Euclidean space  $E$  where the geometric differences between pairs of points in  $E$  agree, as closely as possible, with the true differences between the  $n$  objects.

In our context, the  $n$  objects correspond to the  $N$  probability distributions of each  $M_R$ . Thus, the probability distribution similarities (or dissimilarities) of  $M_R$  can be represented by a  $N \times N$  distance matrix  $\Delta_R = (\delta_{kl})$ , where  $\delta_{kl} = \delta_{lk} \geq 0$  is computed by

$$\delta_{kl} = e^{(K(H_R^{k_l}, H_R^{l_l}))} - 1 \quad (2)$$

where  $K(H_R^{k_l}, H_R^{l_l})$  is the diffusion distance measured between the probability distributions  $k$  and  $l$ , which is proposed in Ling and K. (2006).

Hence, being  $A$  the matrix with entries  $A = -(\frac{1}{2})\delta_{kl}^2$  and the centering matrix  $C = I_n - \frac{1}{n}11'$ , the so-called inner product matrix  $B_R$  associated to  $\Delta_R$  can be computed by  $B_R = CAC$  for each  $M_R$ . The inner product matrix  $B_R$  is  $N \times N$  symmetric matrix which can be spectrally decomposed as  $B_R = U_R \Lambda_R^2 U_R'$ . Assuming the eigenvalues in  $\Lambda_R$  are arranged in descending order, the matrix  $U_R$  represents the standard coordinates of the region  $M_R$  where the  $s$  first columns contain the most relevant region information. It should be remembered that our interest, given two regions defined by  $M_{R_i}$  and  $M_{R_j}$ , is to measure the multivariate association between their  $s$  first standard coordinates. Therefore, a similarity measure is obtained by correlating the principal axis of two region models obtained via MDS. This similarity measure relies on a statistical test based on the multivariate analysis of variance (MANOVA). The goal is to test whether there is a dependence between the principal components of the regions or not. Therefore, two distance matrices  $\Delta_{R_i}$  and  $\Delta_{R_j}$  to find  $B_{R_i} = U_{R_i} \Lambda_{R_i}^2 U_{R_i}'$  and  $B_{R_j} = U_{R_j} \Lambda_{R_j}^2 U_{R_j}'$  should be computed using the explained procedure. The number  $s$  of dimensions is an important aspect in most multivariate analysis methods. In MDS, the number of dimensions is based on the percentage of variability accounted for by the first dimensions. Here, a criterion which extends a sequence  $c$  defined and studied in (Cuadras et al., 2012) is used to set the value of  $s$ . At this point, having two regions defined by their standard coordinates  $U_{R_i}$  and  $U_{R_j}$  whose dimensions are  $N \times s$ , the Wilk's criterion  $W$  for testing  $B=0$  in a multivariate regression model is given by:

$$W(R_i, R_j) = \det(I - U_{R_j}' U_{R_i} U_{R_i}' U_{R_j}) = \prod_{i=1}^s (1 - r_i^2) \quad (3)$$

where  $\det$  means the determinant and  $r_i$  corresponds to the canonical correlation of each axis. Using Eq. 3, the definition of the proposed merging criterion can be defined as:

$$O_{MDS}(R_i, R_j) = \min_{R_i, R_j} W(R_i, R_j) \quad (4)$$

satisfying  $0 \leq W(R_i, R_j) \leq 1$  and  $W(R_i, R_j) = 0$  if  $R_i$  is equal to  $R_j$ .

Once constructed with this region model and similarity measure, the BPT is used as search space for object detection as discussed in the following section. Note that the goal here is not to extract a partition from the BPT and to perform the search on the partition. Instead, all BPT nodes are analyzed.

### 3. Object detection strategy

As instantiations of the object of interest,  $O$ , may have many different visual appearances, the detection relies on a set of features,  $\Omega_F$ , characterizing  $O$ . Based on these features, the likelihood of each BPT node  $P(O|R_i)$  to be an instantiation of  $O$  is assessed and assigned to the node. Once the BPT has been populated with these likelihood, a search is performed to detect the most probable instantiations of the object of interest.

#### 3.1. Populating BPT

For each node  $R_i$ , the likelihood  $P(O|R_i)$  is computed by using a set of spectral and spatial features  $\Omega_F = \{F_1, F_2, \dots, F_K\}$ . Based on these features, the likelihood of each node to be an instantiation of  $O$  can be estimated by the Bayes rule as :

$$P(O|R_i) = P(O|\Omega_F) = \frac{P(\Omega_F|O)P(O)}{P(\Omega_F)} \quad (5)$$

The a priori probability of the object  $P(O)$  is being equally probable to be observed (uniformed prior) and the probability of the evidence  $P(\Omega_F)$  can be viewed as a normalizing constant. Thus, considering independent the  $K$  local features computed at each  $R_i$ , the  $P(O|R_i)$  can be defined by

$$P(O|R_i) \approx P(\Omega_F|O) \approx \prod_{n=1}^K P(F_n|O) \quad (6)$$

The specific choice of features computed on the regions contained at the BPT nodes strongly depends on the reference object. The spatial and spectral features must characterize the shape and the spectral signature of the object. For instance, buildings have a rectangular shape and their spectral signature can be related to asphalt material. In contrast, trees are circular regions having a classical vegetation spectrum. Here, four features are proposed, which leads to the description of the following four feature probabilities:

##### 3.1.1. Spectral features characterizing the object of interest

An hyperspectral scene image is composed of a certain number of spectral classes  $N_c$  defining different types of materials. In general, one specific object can be associated to one material  $C_s$ . For instance, roads can be associated to asphalt material whereas trees can be associated to the vegetation one.

Accordingly, the goal is to compute the spectral class probability distribution  $\mathcal{P}_{R_i}$  (Alonso-Gonzalez et al., 2013) for each BPT node. This distribution describes the probability that the

region  $R_i$  has to belong to all the materials  $N_c$  describing the scene.

Note that this can be done by training a probabilistic SVM pixelwise classifier (Platt et al., 2000) for these classes and using it on the region mean spectrum. As a result, the class probability distribution  $\{\mathcal{P}_{R_i}(C_s)\}_{1 \leq s \leq N_c}$  is available for each node. The  $\mathcal{P}_{R_i}$  computation allows us to define the spectral class probability and the class membership homogeneity features.

*The spectral class probability  $P(F_1|O)$ :* It corresponds to the probability  $\mathcal{P}_{R_i}(C_s)$  that the region  $R_i$  has to belong to the material class  $C_s$  of the object of interest. For instance, for the road detection application, this probability is the likelihood that the region belongs to the asphalt class. This probability is directly extracted from the class probability distribution  $\mathcal{P}_{R_i}$  estimated by the SVM.

*The spectral class membership homogeneity  $P(F_2|O)$ :* This feature evaluates the region homogeneity in terms of class membership. Note that if a region is an object, all its pixels ideally belong to the same class. This term is important in the BPT context, as nodes close to the root node represent regions combining many different classes. It is defined as:

$$P(F_2|R_i) = \sum_{s=1}^{N_c} \sqrt{\mathcal{P}_{R_i}^{R_l}(C_s) \mathcal{P}_{R_i}^{R_r}(C_s)} \quad (7)$$

where  $\mathcal{P}_{R_i}^{R_l}$  and  $\mathcal{P}_{R_i}^{R_r}$  are the class probability distributions of the left and the right child nodes of  $R_i$ . Note that if two sibling nodes have similar class probability distributions, their union will also have a similar distribution, i.e. the object is in the process of being formed.

### 3.1.2. Spatial features characterizing the object of interest

The spatial features of objects are automatically inherited from their structure. In this study, two spatial properties describing the area and the shape of the object have been proposed.

*The region area  $P(F_3|O)$ :* This feature corresponds to the number of pixels forming the region contained in each BPT node. The goal of this feature is to prevent the detection of small or large meaningless regions. It is done by assuming that the area interval  $[\mathcal{A}_{min}, \mathcal{A}_{max}]$  of the object of interest is known.  $P(F_2|O)$  is then defined as a uniform distribution between  $[\mathcal{A}_{min}, \mathcal{A}_{max}]$ . The definition of  $\mathcal{A}_{max}$  is important to detect individual objects as the union of two identical objects can result into a similar object of larger size.

*The area of the smallest oriented bounding box  $P(F_4|O)$ :* This last feature is used to compute a probability related to the region shape. In this work, two different  $P(F_4|O)$  have been used to deal with two different object detection applications. Both are based on the same assumption : the use of a measure normalized between  $[0, 1]$  as a shape probability distribution. In the case of building detection,  $P(F_4|O)$  measures the region

compactness and is the ratio between the area of the region and the area of the smallest oriented bounding box including the region. For road extraction, this term measures the region elongation and it is defined as the ratio between the width and the height of the oriented bounding box.

### 3.2. Processing populated BPT

At this stage, the BPT processing consists in detecting the nodes which are the most likely to be the sought objects. This strategy assumes that the objects of interest appear as individual nodes. The goal is to use the  $P(O|R)$  values to discard nodes that significantly differ from the object of interest and to detect the best object representations. As a first approximation, BPT nodes with high  $P(O|R)$  values are clearly candidates to be the sought objects. At this point, it should be remembered that the BPT structure represents inclusion relationships between regions. As a result, it is likely that nodes belonging to the same tree branch have similar  $P(O|R)$  values than their parent or child nodes. As our goal is to detect non overlapping regions representing instantiations of the object of interest, only the best node  $R^*$  on the branch should be detected.

One solution for the detection of  $R^*$  is to decide that it corresponds to the region with the highest  $P(O|R)$  value. However, this approach based only on the  $P(O|R)$  value is not robust as the maximum can be obtained for nodes close to the leaves where the objects are not yet formed (mainly because  $P(F_1|O)$  and  $P(F_3|O)$  may be high for small regions). Another strategy to detect  $R^*$  is to select the closest node to the root whose  $P(O|R)$  value is higher than a given threshold  $\delta_{Tc}$  (Valero et al., 2011). This approach is somewhat arbitrary since the best node may not always be the closest to the root.

Taking into account these considerations, the approach used here is based on the analysis of the  $P(O|R)$  evolution during the object formation along the branch. If we draw the  $P(O|R)$  values along a BPT branch containing an object of interest starting from the leaf node, the first interesting point of the curve arrives when the smaller regions start having a high  $P(O|R)$  value. After this, a stable range of values where no important change concerning  $P(O|R)$  is generally observed. Finally, the last important step occurs when  $P(O|R)$  suffers an important decrease after a specific merging step. At this point, the resulting region usually corresponds to a non-meaningful object of the image. In these situations, the best object representation  $R^*$  is found just before the important decrease.

An example of this typical evolution can be observed in Fig. 2 where the curve of  $P(O|R)$  values from a leaf to the root is represented.

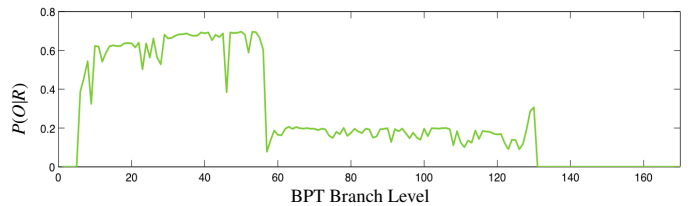


Fig. 2: Example of  $P(O|R)$  evolution along a BPT branch

The horizontal axis indicates the level on the BPT branch. The left side corresponds to the leaf and the right side to the root node) whereas the vertical axis indicates to probability values. In this example, the object formation starts around the fifth BPT level whereas level 58 corresponds to the important decrease where a non-meaningful object is formed. In the case of Fig. 2, the object  $R^*$  is then formed at level 57. We have observed that this behavior is really typical of branches containing the object of interest. Accordingly, the detection of  $R^*$  in a BPT branch is given by

$$R^* = \min_R P(O|R^+) - P(O|R) \text{ , with } P(O|R) > \delta_T \quad (8)$$

where  $R^+$  is the parent node of  $R$  and  $\delta_T$  is the threshold used to decide if a region may be considered as a candidate of the sought object.

As shown in Fig. 3, because of the inclusion relationship described by the BPT, the detection process described above may result in several detections of  $R^*$  along unique BPT branch. In the example of Fig. 3(a), the red and the green branches have been analyzed following Eq.4-8 and two different  $R^*$  are detected depending on the studied branch. Hence, a decision should be taken in order to avoid overlapping regions in the final result. Here, it has been considered that the region analysis is more reliable for large regions. Accordingly, in case of overlap, the  $R^*$  corresponding to the closest region to the root is kept. In the case of Fig. 3(a), the green branch decision is retained as shown in Fig. 3(b). Following this pruning strategy, the selection of the  $R^*$  corresponding to the sought objects is done in a top-down fashion: the BPT is analyzed from the root to the leaves by selecting the first nodes found as  $R^*$ .

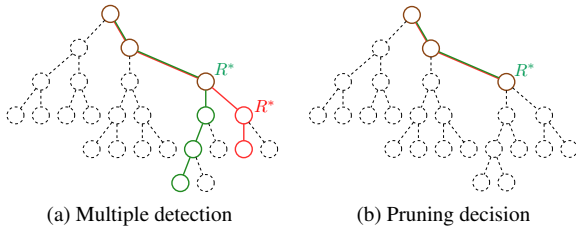


Fig. 3: Multiple detections of  $R^*$  in a same BPT branch

#### 4. Experimental results

This section addresses the evaluation of the object detection strategy proposed in Section 3. The goal of the experiments is to compare the results of the proposed strategy with a classical pixel-wise method such as SVM classification. In order to perform this evaluation, detection examples of two different urban objects: roads and buildings, are discussed. The experimental evaluation is carried out using two different hyperspectral images captured by two different sensors.

The first studied hyperspectral image was acquired over Pavia (Italy) by the ROSIS sensor having a 1.3m spatial resolution. It corresponds to a urban area and the hyperspectral

data involves 102 spectral bands. The ground truth (available at [http://www.ehu.es/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes&redirect=no](http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes&redirect=no)) is composed of 9 classes and 7456 samples

The experiment targets the detection of buildings. The evaluation is performed on two different portions of the complete image shown in Fig. 4 as RGB false-color compositions of three hyperspectral bands. On these images, the BPTs are computed with the procedure described in Section 2.



Fig. 4: False color composition of two portions of the Pavia urban hyperspectral data used for building detection

Once the BPT has been computed, the four features presented in Section 3 are computed. The class probability distribution  $\{\mathcal{P}_{R_i}(c_s)\}_{1 \leq s \leq N_c}$  is estimated with a SVM Gaussian kernel function constructed through a training step. This step follows the classical cross-validation strategy: the training set is divided into  $k$  parts, the SVM is trained using  $(k-1)$  parts and the obtained parameters are tested on the remaining part. The SVM training step is done by selecting randomly 20% of samples for each class from the available reference data. Once the kernel function is constructed, it is used to assign to each BPT node their class probability distribution  $\{\mathcal{P}_{R_i}(c_s)\}$ . In order to classify nodes corresponding to regions with several pixels, the region mean spectrum is used as input to the SVM classifier.

The constructed SVM kernel function is also used to perform a pixel-wise classification. The corresponding classification results obtained for the class of buildings are shown in Fig. 5. As can be seen, these classification results are rather noisy.

The results obtained by the proposed BPT-based strategy are shown in Fig. 5. In this case, the  $\delta_T$  parameter is set to 0.65 and the range  $[\mathcal{A}_{min}, \mathcal{A}_{max}]$  is set to  $[30, 1000]$ . As can be seen, most of the rectangular buildings have been precisely detected. These results corroborate the advantage of using the BPT representation. The use of spectral as well as spatial descriptors of BPT nodes clearly outperforms the classical pixel-wise detection using only spectral information. On the other hand, it should be also remarked that the results shown in Fig. 5 are also comparable with the results obtained by (Akay and Aksoy, 2008), where a building detection map is also presented on the same hyperspectral Pavia image.

The second experiment has been performed using two portions of a publicly available HYDICE hyperspectral data (available at <http://www.agc.army.mil/hypercube/>). After removing water absorption and noisy bands, the data contain 167 spectral bands shown in Fig. 6 as a RGB combination of three bands. The data set is composed of 8 classes and 4712 ground truth samples. The same building detection experiment has been carried out on this dataset. In this case, the range  $[\mathcal{A}_{min}, \mathcal{A}_{max}]$

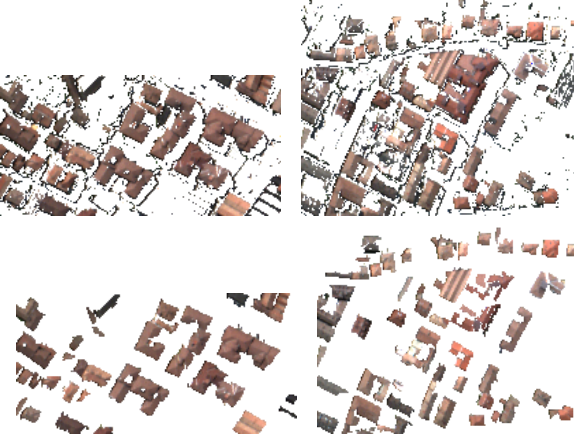


Fig. 5: Obtained results on Pavia urban data. First row: Pixel-wise SVM classification. Second row: BPT-based detection

is set to  $[10, 400]$  due to the lower spatial resolution of this image (approximately 3m).

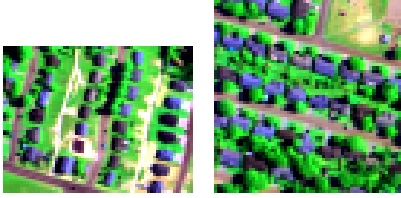


Fig. 6: RGB combination of Hydice urban scene

The obtained results are shown in Fig. 7 where the first column shows the manually created ground truth. The obtained results for the SVM pixel-wise classification and BPT strategy are shown in Fig. 7(b) and Fig. 7(c). The SVM results have been obtained by repeating 10 times the random selection of the training data set. Looking at these results, the benefit of incorporating the spatial information by using the BPT representation is also remarkable. Thanks to the availability of ground truth, the results of Fig. 7 are also evaluated objectively by measuring the *precision* and *recall* values. The precision corresponds to the percentage of positive detection that are correct, whereas the recall indicates how many pixels belonging to the object are correctly identified as such.

These measures are computed from the number of true positives  $TP$  (pixels correctly labeled), of false positives  $FP$  (pixels incorrectly labeled as building class) and of false negatives  $FN$  (pixels not labeled as building but actually belonging to a building). The obtained precision and recall measures are reported in Table 1. The recall evaluation shows how the two methods detect most of the pixels belonging to the buildings. However, the evaluation in terms of precision corroborates the good performance of the BPT strategy against the pixel-wise detection.

At this point, the importance of the area feature  $\mathcal{A}_{max}$  and the  $\delta_T$  may be discussed. Let us consider the small portion of Fig. 7(a) located at the lower left corner of the bottom image. A zoom of this area can be observed at Fig. 8(a) where two buildings appear very close. For this example, the detected BPT

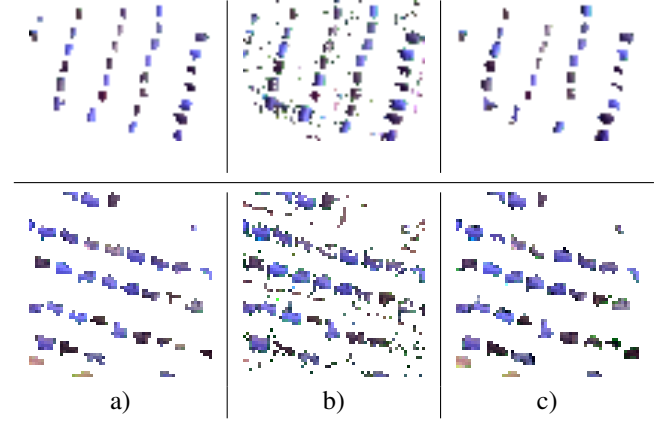


Fig. 7: Building detection example on Hydice urban scene. a) Ground truth, b) Pixel-wise classification, c) BPT-based detection

Table 1: Quantitative evaluation for building detection

	$\text{Precision} = \frac{TP}{(TP+FP)}$		$\text{Recall} = \frac{TP}{(TP+FN)}$	
	SVM pixel-wise	BPT	SVM pixel-wise	BPT
Fig.7 (top)	0.47	0.73	0.91	0.89
Fig.7 (bottom)	0.57	0.74	0.83	0.88

node at Fig. 7(c) is shown in Fig. 8(c). As it can be observed, the two buildings have been detected as a unique object. Therefore, the maximum decrease of the curve does not correspond to the best node representing the building.

This fact can be explained by studying the  $P(O|R)$  evolution from a pixel belonging to the left building. The resulting curve is plotted at Fig. 8(d) where two  $\delta_T$  values are highlighted in red and green.

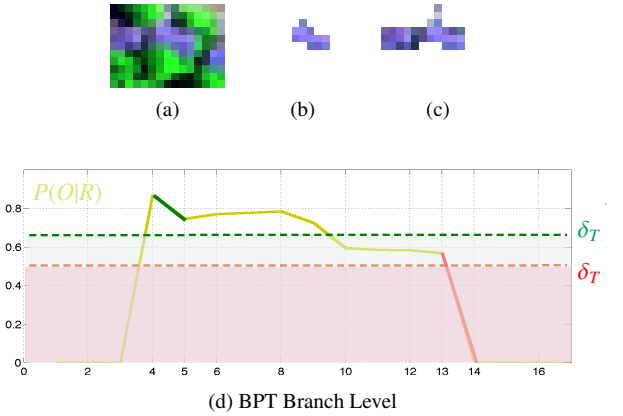


Fig. 8: (a) False color composition of a small area of Fig. 6, (b) BPT node at level 4, (c) BPT node at level 13,  $P(O|R)$  evolution along a BPT branch.

Looking at this figure, it is observed that the detected node shown in Fig. 8(c) corresponds to the branch level 13 by using  $\delta_T = 0.5$ . However, the merging at the level 5 should not be done since the best representation appears at level 4. However, it must be remarked that the resulting region is indeed a candidate according to the  $\mathcal{A}_{max}$  feature. In fact,  $\mathcal{A}_{max}$  should be used here to detect that the resulting region corresponds to the

union of two objects. However, a building having an area equal to  $\mathcal{A}_{max}$  may actually exist. Besides, the example of Fig. 8(d) shows the importance of  $\delta_T$  definition. As it can be observed, the  $\delta_T = 0.65$  detects the region at level 4 as  $R^*$  whereas  $\delta_T = 0.5$  has detected the level 13. Hence, this example shows how the definition of  $\mathcal{A}_{max}$  and  $\delta_T$  may not be straightforward and it must be chosen as a compromise. To demonstrate the flexibility of the BPT approach, a second example aiming at road extraction is proposed. In this case, the same HYDICE images presented on Fig. 6 are used. For road extraction, the first three features previously computed for the building application are used. However, the region elongation has been computed as the fourth feature describing the shape of the object. As mentioned above, the elongation of region is the ratio between the width and the height of the oriented bounding box. This measure ranges from 0 to 1 and is used as  $P(F_4|O)$ . The obtained results are shown in Fig. 9(b)(c).

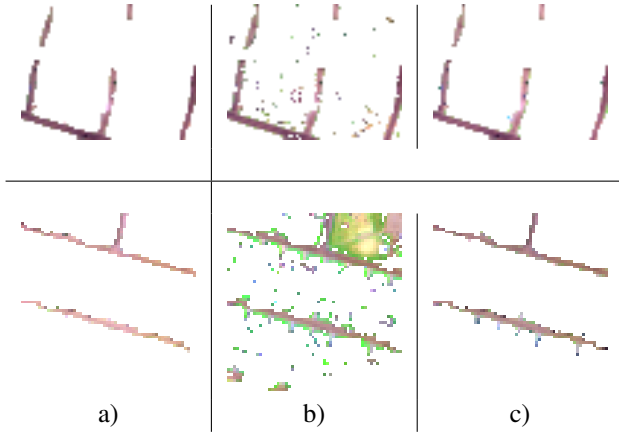


Fig. 9: Road detection example on Hydice urban scene. a) Ground truth, b) Pixel-wise classification, c) BPT-based detection

The visual evaluation clearly shows how roads do not appear only as pixels whose radiance values are similar to asphalt and the improvement provided by the BPT approach is also quite significant. A quantitative evaluation based on precision and recall has also been carried out by using the ground truth shown in Fig.9(a). Tab.2 shows the precision and recall values. Looking at these results, it can be observed how the BPT approach also obtains the better results in this experiment, in particular for the precision values.

Table 2: Quantitative evaluation for road detection

	Precision = $\frac{TP}{(TP+FP)}$		Recall = $\frac{TP}{(TP+FN)}$	
	SVM pixel-wise	BPT	SVM pixel-wise	BPT
Fig.9 (top)	0.669	0.875	0.9327	0.9432
Fig.9 (bottom)	0.223	0.7344	0.8814	0.9289

## 5. Conclusions

An automatic hyperspectral object detection methodology using a BPT image representation has been detailed in this

work. It has been illustrated how BPT can be a powerful image representation which provides a hierarchically structured search space for object recognition applications where the spectral and the spatial information can be incorporated in the search of a reference object. The obtained results show the interest of studying the objects of the scene with a region-based perspective and to avoid reducing the search space by producing a partition as preprocessing. [This new object-based analysis can open the door to an important number of techniques exploiting the extremely high resolution \(very few centimeters\) imagery such as hyperspectral UAV images.](#) Future works will be conducted on the detection of other urban structures using the presented methodology.

## References

- Akay, H., Aksoy, S., 2008. Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE Transactions on Geoscience and Remote Sensing* 46 (7), 2097–2111.
- Aksoy, S., Younan, N. H., Bruzzone, L., 2010. Editorial: Pattern recognition in remote sensing. *Pattern Recognition Letters* 31 (10), 1069–1070.
- Alonso-Gonzalez, A., Valero, S., Chanussot, J., Lopez-Martinez, C., Salembier, P., 2013. Processing multidimensional sar and hyperspectral images with binary partition tree. *IEEE Transactions on Geoscience and Remote Sensing* 101, 723 – 747.
- Aytekin, O., Ulusoy, I., 2011. Automatic segmentation of vhr images using type information of local structures acquired by mathematical morphology. *Pattern Recognition Letters* 32 (13), 1618–1625.
- Bernstein, R., Gesu, V. D., 1999. A combined analysis to extract objects in remote sensing images. *Pattern Recognition Letters* 20 (11), 1407 – 1414.
- Calderero, F., Marqués, F., 2010. Region-merging techniques using information theory statistical measures. *IEEE Transactions on Image Processing* 19(6), 1567 – 1586.
- Cox, T., Cox, M., 1994. *Multidimensional Scaling*. K. Fernandez and A. Morineau (Ed.), Chapman & Hal.
- Cracknell, A., 1998. Synergy in remote sensing. what’s in a pixel? *International Journal of Remote Sensing* 19 (11), 2025–2047.
- Cuadras, C., Valero, S., Cuadras, D., Salembier, P., Chanussot, J., 2012. Distance-based measures of association with applications in relating hyperspectral images. *Communications in Statistics - Theory and Method* 41, 2342–2355.
- Darwish, A., Leukert, K., Reinhardt, W., 2003. Image segmentation for the purpose of object-based classification, 2039–2041.
- Dimiccoli, M., Salembier, P., 2009. Hierarchical region-based representation for segmentation and filtering with depth in single images. *IEEE Proceedings of ICIP*.
- Fauvel, M., Chanussot, J., Benediktsson, J., 2012. A spatial-spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognition Letters* 45 (1), 381 – 392.
- Ling, H., K., O., 2006. Diffusion distance for histogram comparison.
- Platt, J., Smola, A., Bartlett, P., Scholkopf, B., Schuurman, D., 2000. Probabilities for support vector machines. in *Advances in Large Margin Classifiers* Eds. Cambridge MA: MIT Press, 61 – 74.
- Plaza, A., Tilton, J., 2005. Automated selection of results in hierarchical segmentations of remotely sensed hyperspectral images., 4946–4949.
- Tarabalka, Y., Chanussot, J., Benediktsson, J., 2010. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition Letters* 43 (7), 2367 – 2379.
- Valero, S., Chanussot, J., Benediktsson, J., Talbot, H., Waske, B., 2010. Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images. *Pattern Recognition Letters* 31 (10), 1120 – 1127.
- Valero, S., Salembier, P., Chanussot, J., 2013. Hyperspectral image representation and processing with binary partition trees. *IEEE Transactions on Image Processing* 22, 1430 – 1443.
- Valero, S., Salembier, P., Chanussot, J., Cuadras, C., 2011. Improved binary partition tree construction for hyperspectral images: Application to object detection. *IEEE Proceedings of IGARSS*, 2515 – 2518.